BIS2C Winter 2026 (Ward)

Lecture 3. Estimating trees

Today's first lecture:

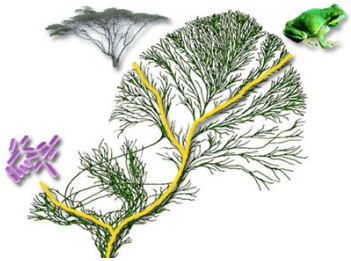
- **Film: "Discovering the Great Tree of Life"**
- **Estimating trees with parsimony**

Second lecture

- Estimating trees (cont'd)
- Model-based approaches
- Interpreting trees

BIS2C Winter 2026 (Ward)

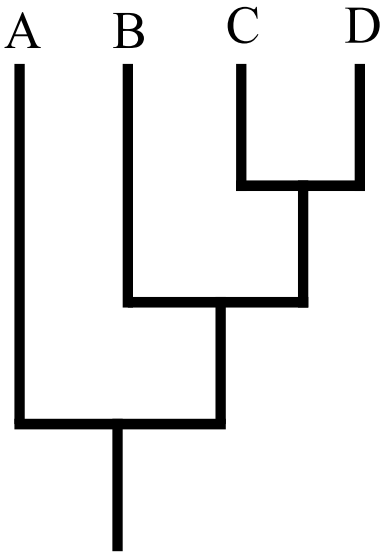
Lecture 3. Estimating trees



Learning Goals. Be able to:

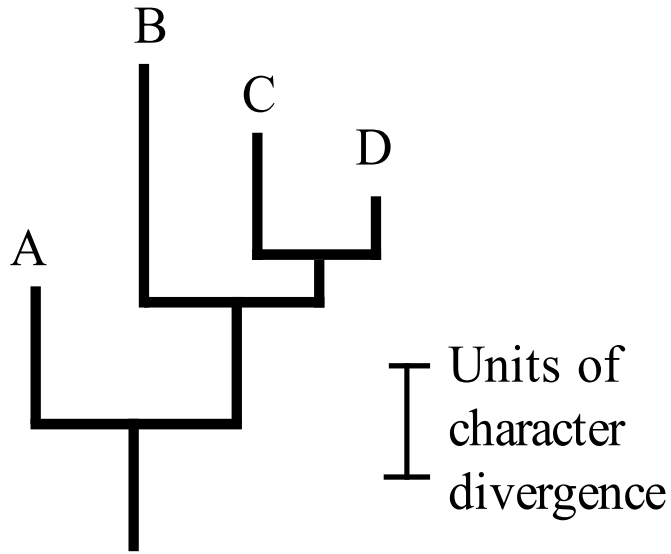
1. Describe how a **taxon/character matrix** is assembled.
2. Distinguish between **rooted** and **unrooted** trees.
3. Explain how a tree is estimated under the principle of **parsimony**.
4. **Estimate the most parsimonious tree** by performing calculations on a simple taxon/character matrix.

Different kinds of trees



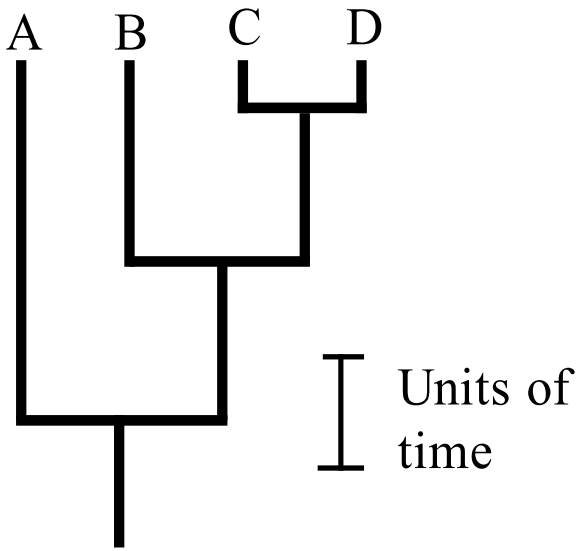
Cladogram

Only the **relative branching order** is depicted. No meaning to branch lengths.



Phylogram

Branch length is **proportional to the amount of character change**



Chronogram

Branch length is **proportional to time**

Recap: Synapomorphy and Homology

Synapomorphy (of a clade): a trait that evolved in the ***most recent common ancestor*** (MRCA) of the clade.

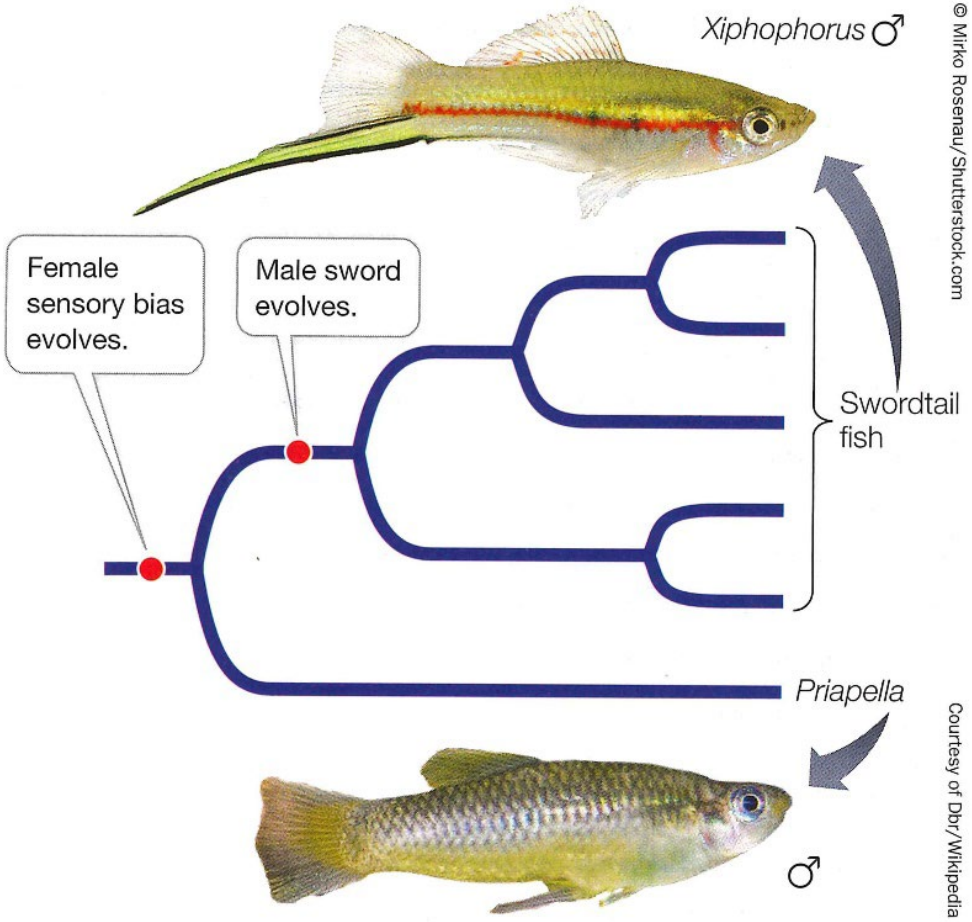
Homology: similarity due to inheritance from a common ancestor.

For a specified clade: all synapomorphies are homologies, but not all homologies are synapomorphies. Why? Because some homologous traits evolved in a **more distant ancestor** than the most recent common ancestor.

Fig. 20.9 in your textbook depicts a phylogeny of swordtail fish.

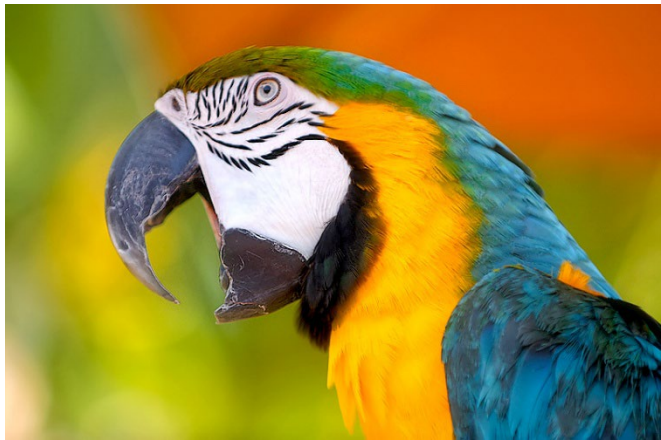
Male sword **is** a synapomorphy of swordtails.
Female sensory bias **is not** a synapomorphy of swordtails.

They are both homologous traits within the swordtail clade.



Life, 12th edition
Fig. 20.9

Some organisms are **more closely related** to one another than to others, i.e., **they share a more recent common ancestor**.



How do we estimate phylogenies?

Phylogenetic trees are *inferred* from the **similarities** that are observed among species or populations (terminal taxa). A tree is an **hypothesis** that can be tested with additional data.

Traits or features of organisms are referred to as **characters**. To infer phylogenies we need characters that exhibit some variation, i.e., characters that show different **character states**.

Examples of characters with **discrete** character states:

Character	**Character states**
1. Wings	present, absent
2. Eye color	red, white, brown
3. Leg length	short, long
4. No. of antennal segments	11, 12
5. Nucleotide site (DNA)	adenine, thymine, guanine, cytosine

How do we estimate phylogenies?

For ease of data management, we can apply **codes** to the character states.

Character	**Character states**	**Codes**
1. Wings	present, absent	0, 1
2. Eye color	red, white, brown	0, 1, 2
3. Leg length	short, long	0, 1
4. No. of antennal segments	11, 12	0, 1
5. Nucleotide site (DNA)	adenine, thymine…	A, T, G, C

How do we estimate phylogenies?

Next, we need to record the character states for each of the terminal taxa that we are studying. For this purpose, we prepare a **data matrix** (taxon-by-character matrix).

	Character									
	1	2	3	4	5	6	7	8	9	10…
Taxon										
Bat	1	0	1	0	0	0	A	C	C	T…
Rat	0	1	1	0	0	1	G	C	C	T…
Cat	0	1	0	1	1	2	G	T	C	T…
Dog	0	1	0	1	1	2	G	T	T	T…
	morphology						DNA sequences			

Now we need to **evaluate the similarity** between taxa and create a tree.

	Character									
	1	2	3	4	5	6	7	8	9	10…
Taxon										
Bat	1	0	1	0	0	0	A	C	C	T…
Rat	0	1	1	0	0	1	G	C	C	T…
Cat	0	1	0	1	1	2	G	T	C	T…
Dog	0	1	0	1	1	2	G	T	T	T…

If we apply the principle of **parsimony** then we search for a tree that requires the **least amount of change in character states** (the fewest "steps" where a step is a change from one character state to another).

	Character									
	1	2	3	4	5	6	7	8	9	10…
Taxon										
Bat	1	0	1	0	0	0	A	C	C	T…
Rat	0	1	1	0	0	1	G	C	C	T…
Cat	0	1	0	1	1	2	G	T	C	T…
Dog	0	1	0	1	1	2	G	T	T	T…

We call this tree with the fewest steps (least amount of character state change) the **most parsimonious tree (or mpt)**.

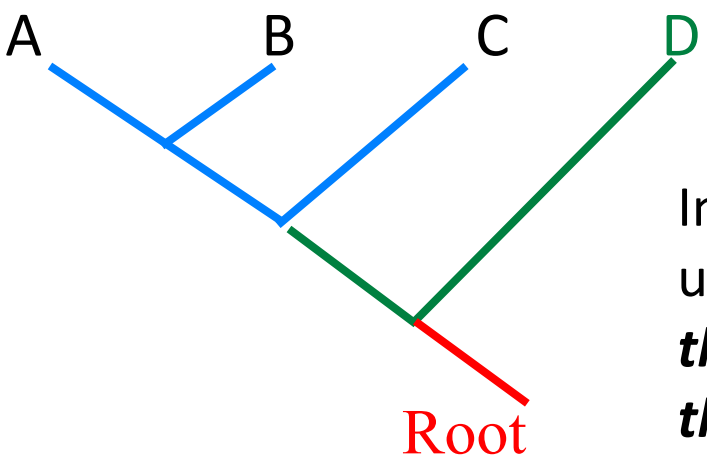
	Character									
	1	2	3	4	5	6	7	8	9	10…
Taxon										
Bat	1	0	1	0	0	0	A	C	C	T…
Rat	0	1	1	0	0	1	G	C	C	T…
Cat	0	1	0	1	1	2	G	T	C	T…
Dog	0	1	0	1	1	2	G	T	T	T…

In the lab you will first learn about reconstructing the most parsimonious tree using the **“outgroup method”**. Here you use one or more outgroups to estimate the ancestral character states (the character states in the MRCA of the ingroup).

In the lecture we start with a second method for inferring phylogenies (introduced later in the lab) that is actually the way most tree reconstruction algorithms (programs) operate. This method involves working with **unrooted trees**.

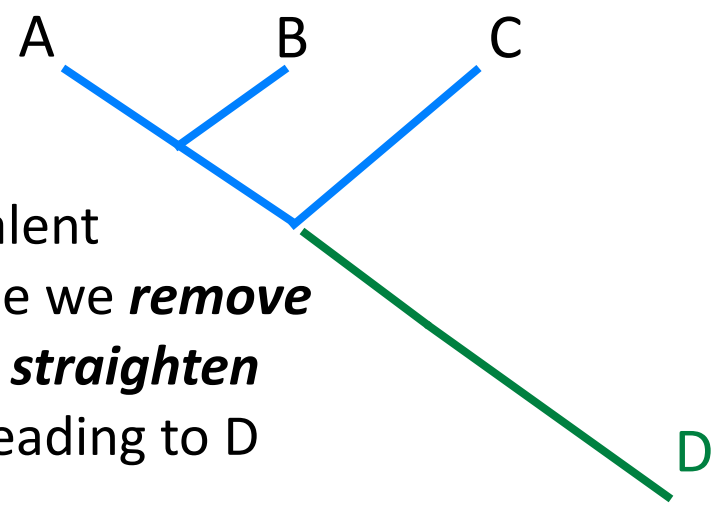
What is an **unrooted tree**? It is a tree in which the root has been removed and relationships among taxa are more ambiguous.

Rooted tree

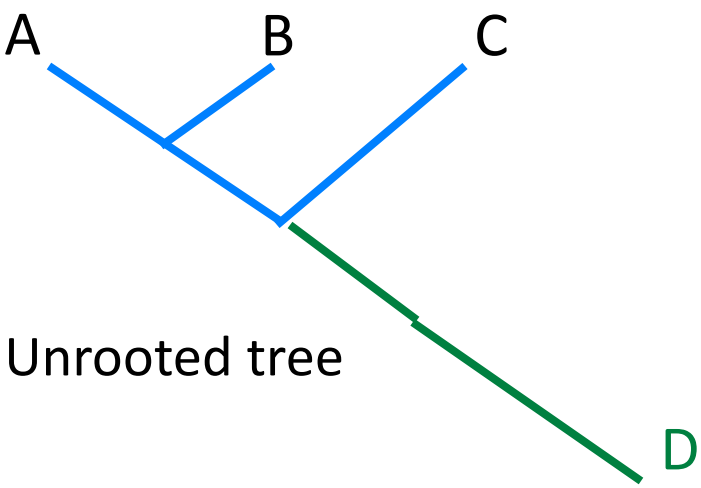


In the equivalent unrooted tree we ***remove the root*** and ***straighten the branch*** leading to D

Unrooted tree



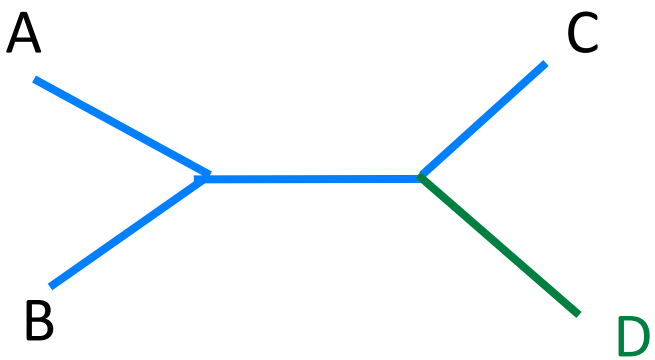
An unrooted tree shows branch splitting events, but it says nothing about the *temporal sequence* of these events. It is a tree without a sense of direction.



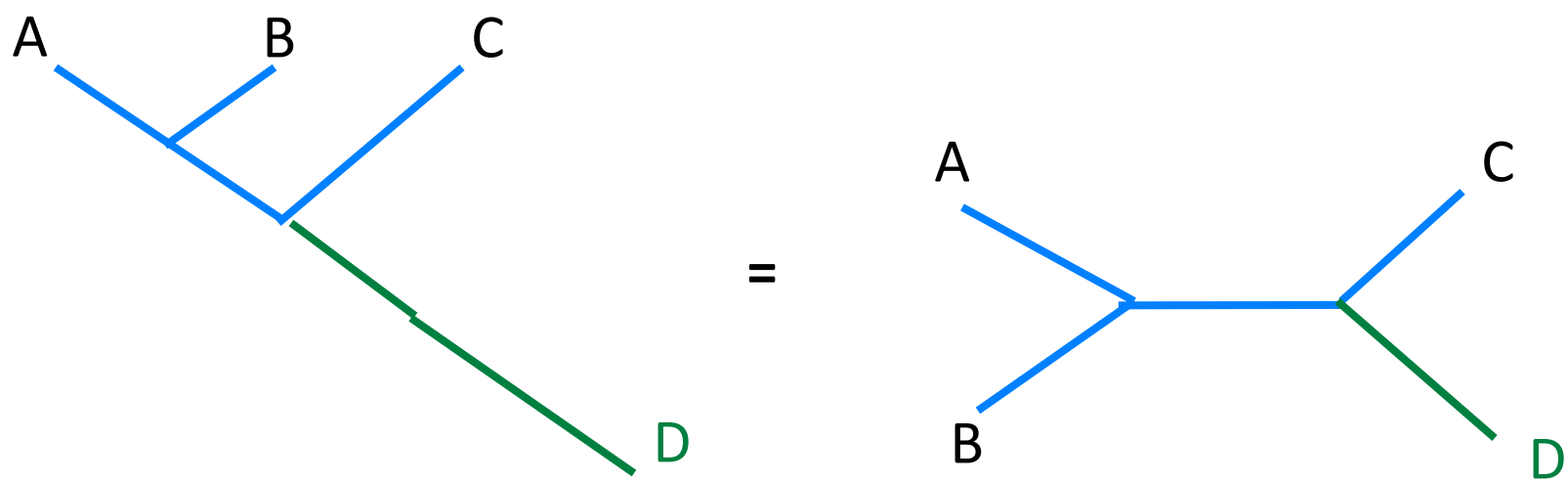
Unrooted tree

=

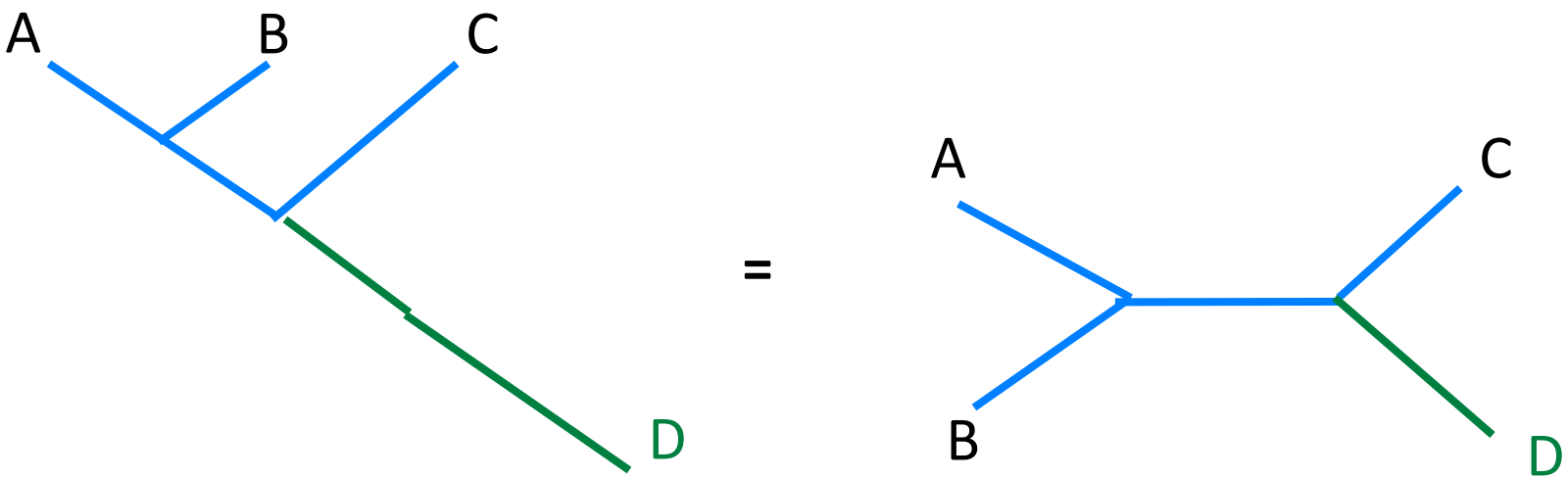
This is the same unrooted tree as the one on the left



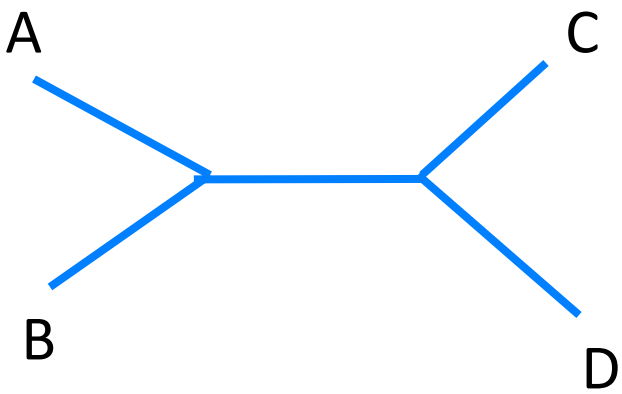
In most tree reconstruction methods we first search for the “best” unrooted tree, and then at the very end we identify an outgroup (if possible) and root the tree with this outgroup.



What is the “best” unrooted tree? If we are using parsimony then it will be the tree which requires the least amount of character state change (fewest steps).



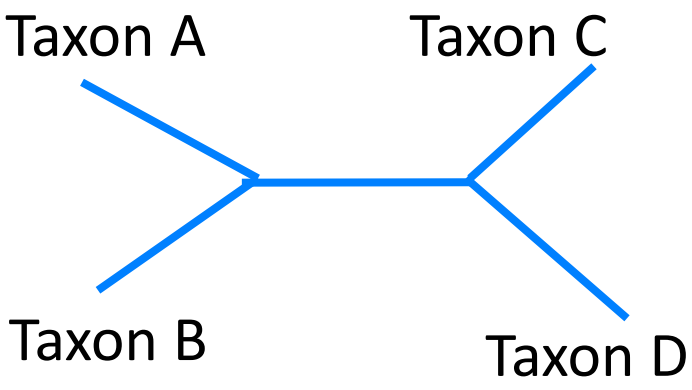
How many possible unrooted trees are there for four taxa?



Clicker Question #1

How many possible unrooted trees are there for four taxa?

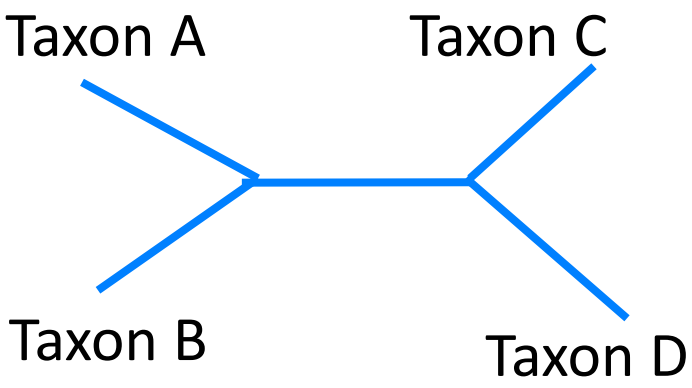
A. 1 tree
B. 2 trees
C. 3 trees
D. 4 trees
E. 8 trees



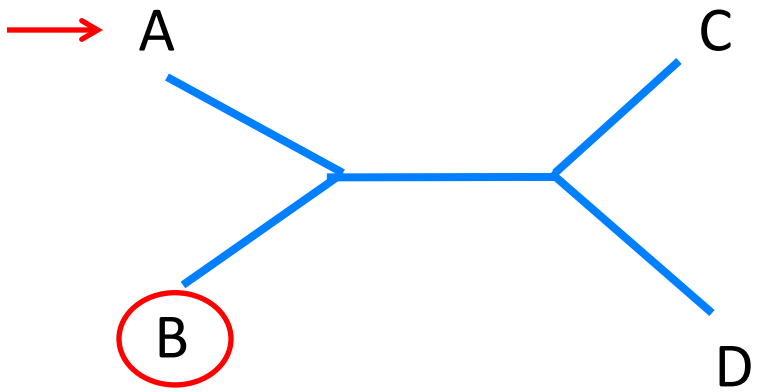
Clicker Question #1

How many possible unrooted trees are there for four taxa?

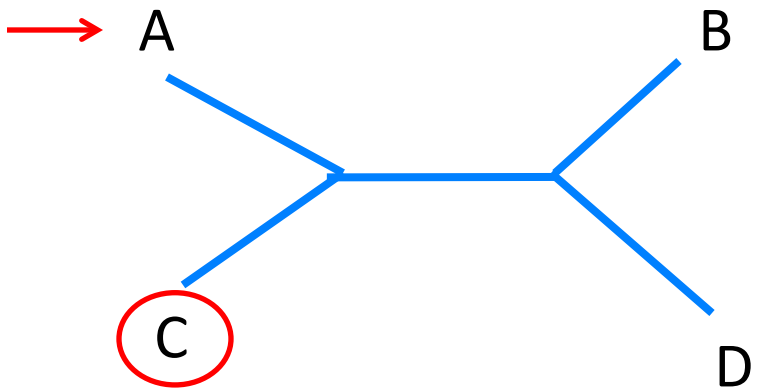
A. 1 tree
B. 2 trees
→ C. 3 trees
D. 4 trees
E. 8 trees



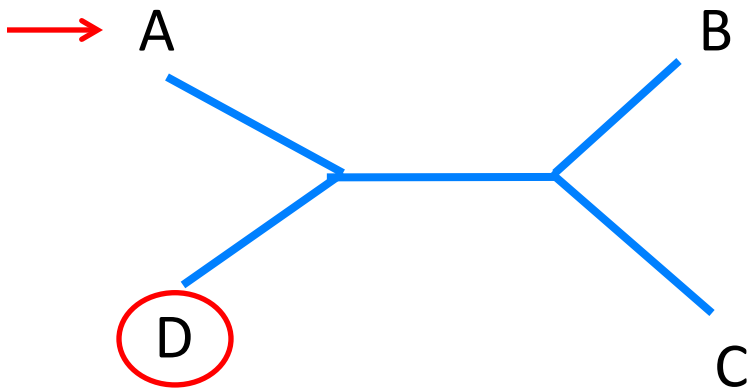
How many possible unrooted trees are there for four taxa?



How many possible unrooted trees are there for four taxa?



How many possible unrooted trees are there for four taxa?



Answer: Only three

So, let's try resolving the phylogeny of four taxa

	Character								
	1	2	3	4	5	6	7	8	9
Taxon									
Mantid	C	T	G	G	A	C	C	C	T
Wasp	C	T	A	G	C	T	C	C	C
Beetle	C	T	G	G	C	C	A	C	T
Fly	T	T	A	G	C	T	A	C	T

Basic procedure for finding/searching for the most parsimonious tree.

1. Ignore parsimony-uninformative characters (for the moment).
2. Draw possible trees.
3. For each informative character count the minimum number of steps required on each tree.
4. Sum over all characters to get the "score" (number of steps) for each tree. (Add the counts for parsimony-uninformative characters.)
5. The most parsimonious tree(s) is/are the one(s) with lowest score.
6. Root the tree along the branch leading to the outgroup (if known).

Which characters in this data matrix are **parsimony-uninformative**?

	Character								
	1	2	3	4	5	6	7	8	9
Taxon									
Mantid	C	T	G	G	A	C	C	C	T
Wasp	C	T	A	G	C	T	C	C	C
Beetle	C	T	G	G	C	C	A	C	T
Fly	T	T	A	G	C	T	A	C	T

Which characters in this data matrix are parsimony-uninformative?

(1) Invariant characters

	Character								
	1	2	3	4	5	6	7	8	9
Taxon									
Mantid	C	T	G	G	A	C	C	C	T
Wasp	C	T	A	G	C	T	C	C	C
Beetle	C	T	G	G	C	C	A	C	T
Fly	T	T	A	G	C	T	A	C	T

Which characters in this data matrix are parsimony-uninformative?

(1) Invariant characters (2) One taxon different, others all the same

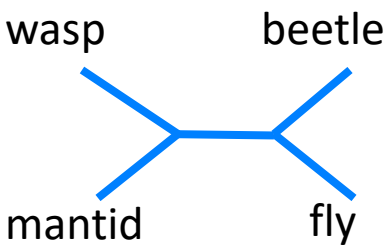
	Character								
	1	2	3	4	5	6	7	8	9
Taxon									
Mantid	C	T	G	G	A	C	C	C	T
Wasp	C	T	A	G	C	T	C	C	C
Beetle	C	T	G	G	C	C	A	C	T
Fly	T	T	A	G	C	T	A	C	T

Which characters are **parsimony-informative**? In the case of four taxa, only those characters where *two taxa have one state, and two taxa have a second state.*

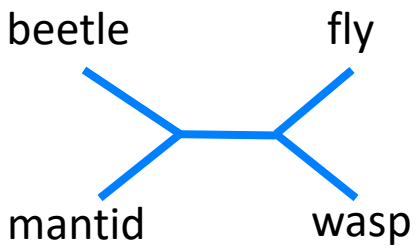
	Character								
	1	2	**3**	4	5	**6**	**7**	8	9
Taxon									
Mantid	C	T	**G**	G	A	**C**	**C**	C	T
Wasp	C	T	**A**	G	C	**T**	**C**	C	C
Beetle	C	T	**G**	G	C	**C**	**A**	C	T
Fly	T	T	**A**	G	C	**T**	**A**	C	T

Number of steps required for each character on each tree

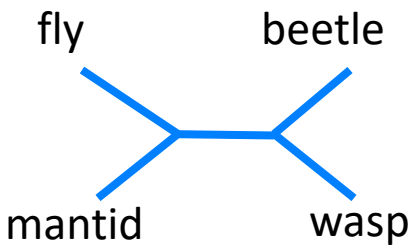
Tree #1



Tree #2



Tree #3

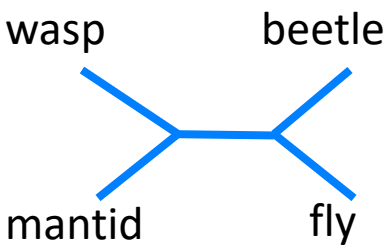


	Tree #1	Tree #2	Tree #3
Char 3	2	1	2
Char 6	2	1	2
Char 7	1	2	2

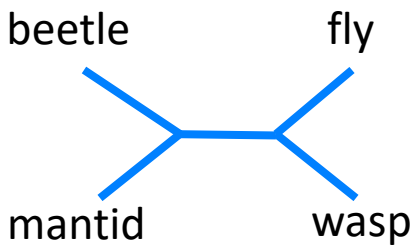
	Character								
	1	2	**3**	4	5	**6**	**7**	8	9
Taxon									
Mantid	C	T	**G**	G	A	**C**	**C**	C	T
Wasp	C	T	**A**	G	C	**T**	**C**	C	C
Beetle	C	T	**G**	G	C	**C**	**A**	C	T
Fly	T	T	**A**	G	C	**T**	**A**	C	T

Number of steps required for each character on each tree

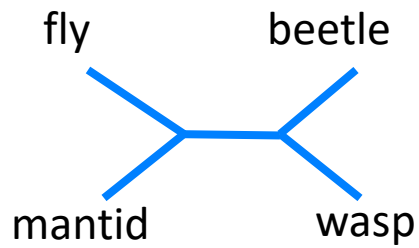
Tree #1



Tree #2



Tree #3

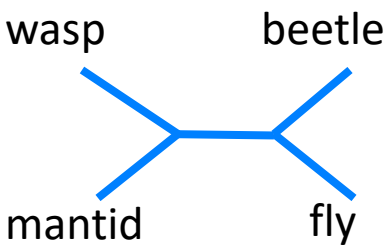


Char 3	2	1	2
Char 6	2	1	2
Char 7	1	2	2
Total	**5**	**4**	**6**

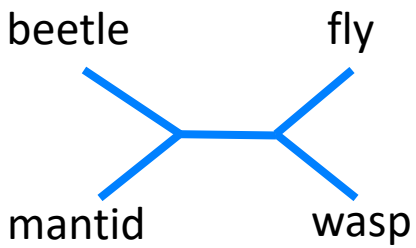
	Character								
	1	2	**3**	4	5	**6**	**7**	8	9
Taxon									
Mantid	C	T	**G**	G	A	**C**	**C**	C	T
Wasp	C	T	**A**	G	C	**T**	**C**	C	C
Beetle	C	T	**G**	G	C	**C**	**A**	C	T
Fly	T	T	**A**	G	C	**T**	**A**	C	T

Number of steps required for each character on each tree

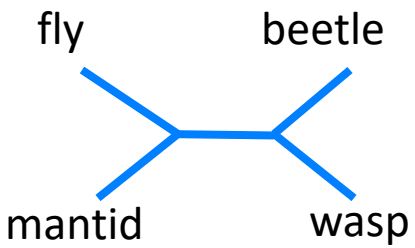
Tree #1



Tree #2



Tree #3



Char 3	2	1	2
Char 6	2	1	2
Char 7	1	2	2
Total	**5 + 3**	**4 + 3**	**6 + 3**

Plus three additional steps for each of the three parsimony-uninformative characters

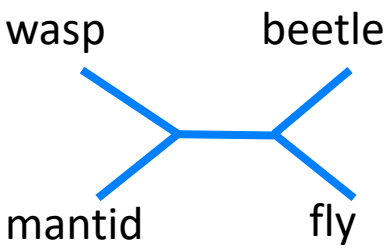
Which characters in this data matrix are parsimony-uninformative?

(1) Invariant characters (2) One taxon different, others all the same

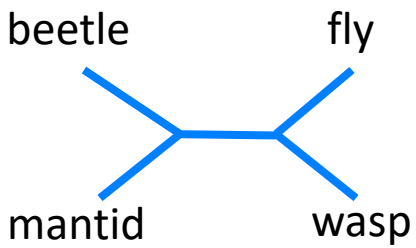
	Character								
	1	2	3	4	5	6	7	8	9
Taxon									
Mantid	C	T	G	G	A	C	C	C	T
Wasp	C	T	A	G	C	T	C	C	C
Beetle	C	T	G	G	C	C	A	C	T
Fly	T	T	A	G	C	T	A	C	T

Number of steps required for each character on each tree

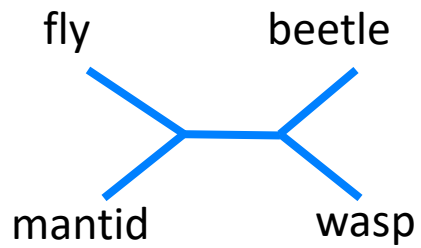
Tree #1



Tree #2



Tree #3

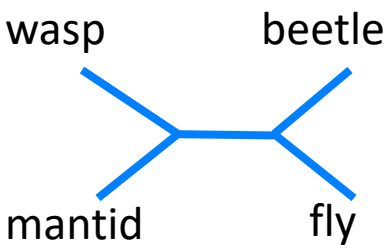


Char 3	2	1	2
Char 6	2	1	2
Char 7	1	2	2
Total	**5 + 3 = 8**	**4 + 3 = 7**	**6 + 3 = 9**

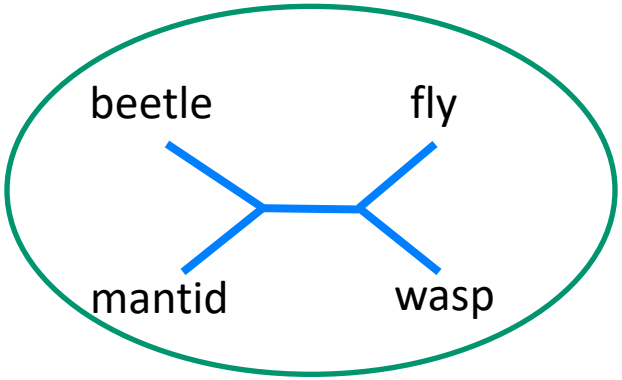
Plus three additional steps for each of the three parsimony-uninformative characters

Number of steps required for each character on each tree

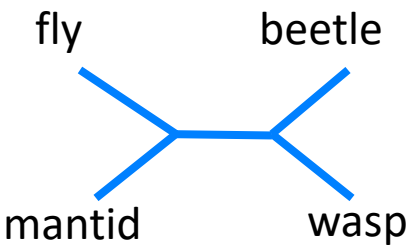
Tree #1



Tree #2



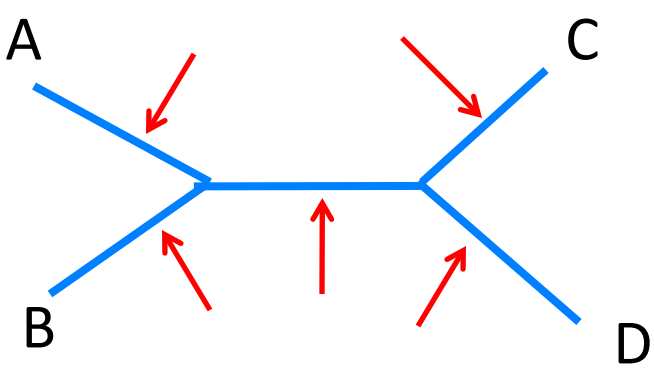
Tree #3



Char 3	2	1	2
Char 6	2	1	2
Char 7	1	2	2
Total	**5 + 3 = 8**	**4 + 3 = 7**	**6 + 3 = 9**

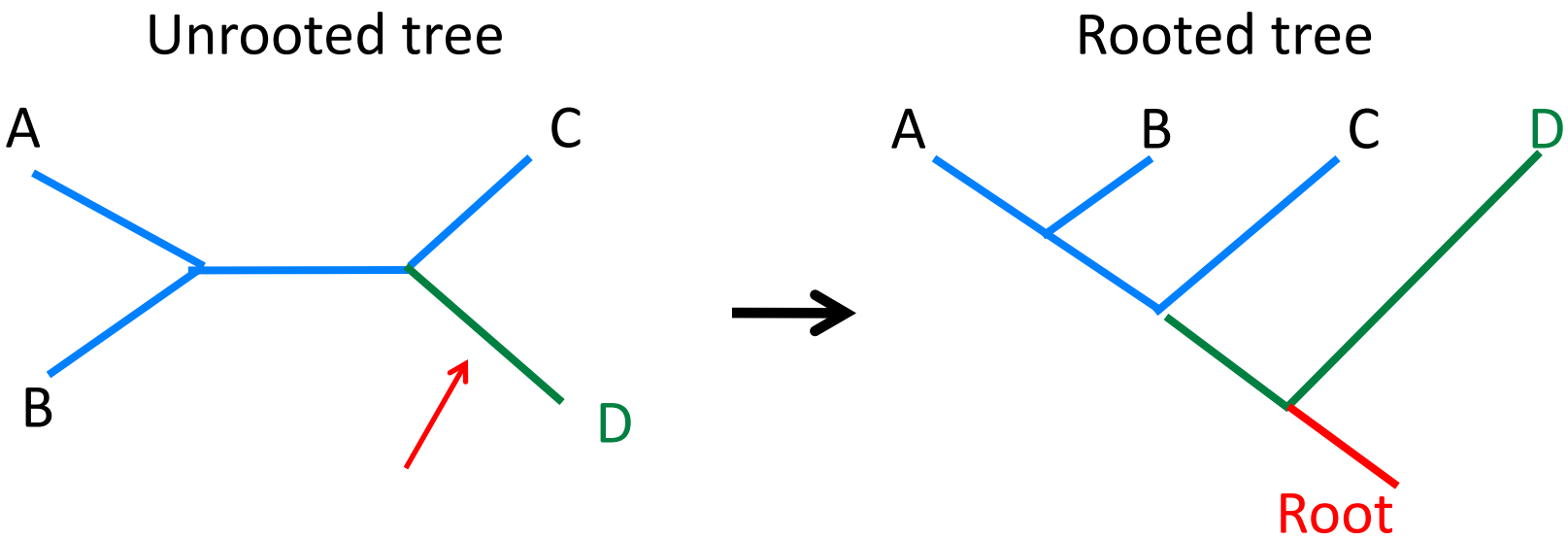
Plus three additional steps for each of the three parsimony-uninformative characters

How do we root an unrooted tree?

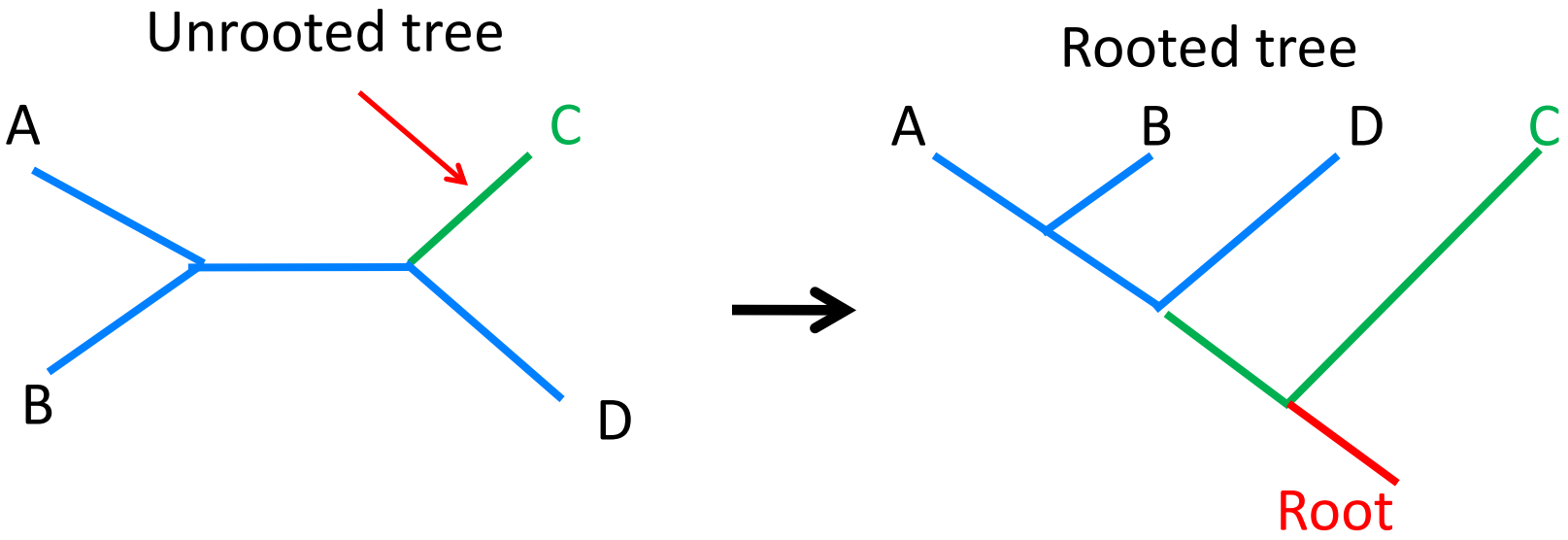


This unrooted tree of four taxa can be converted into **five** different rooted trees, depending on where the root is attached.

If D is the outgroup, then we attach the root on the branch leading to D.

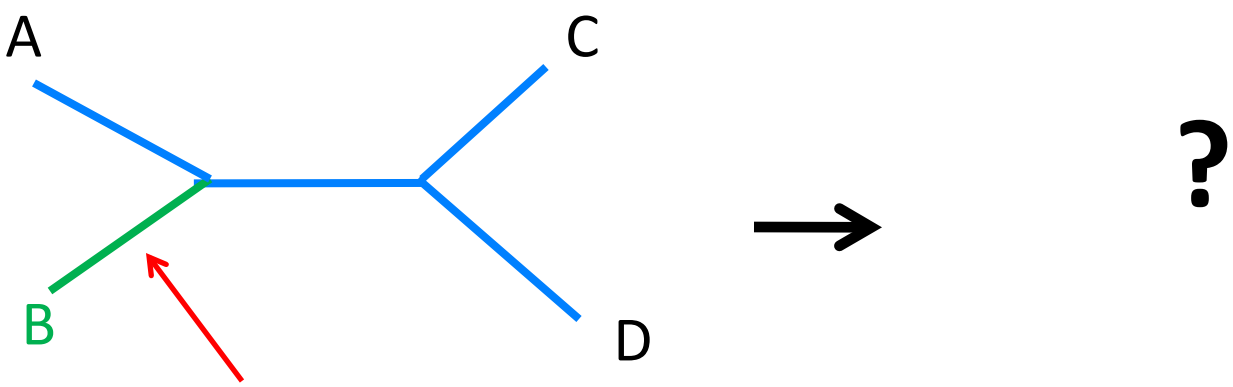


If C is the outgroup, then we attach the root on the branch leading to C and **we get a different rooted tree**.

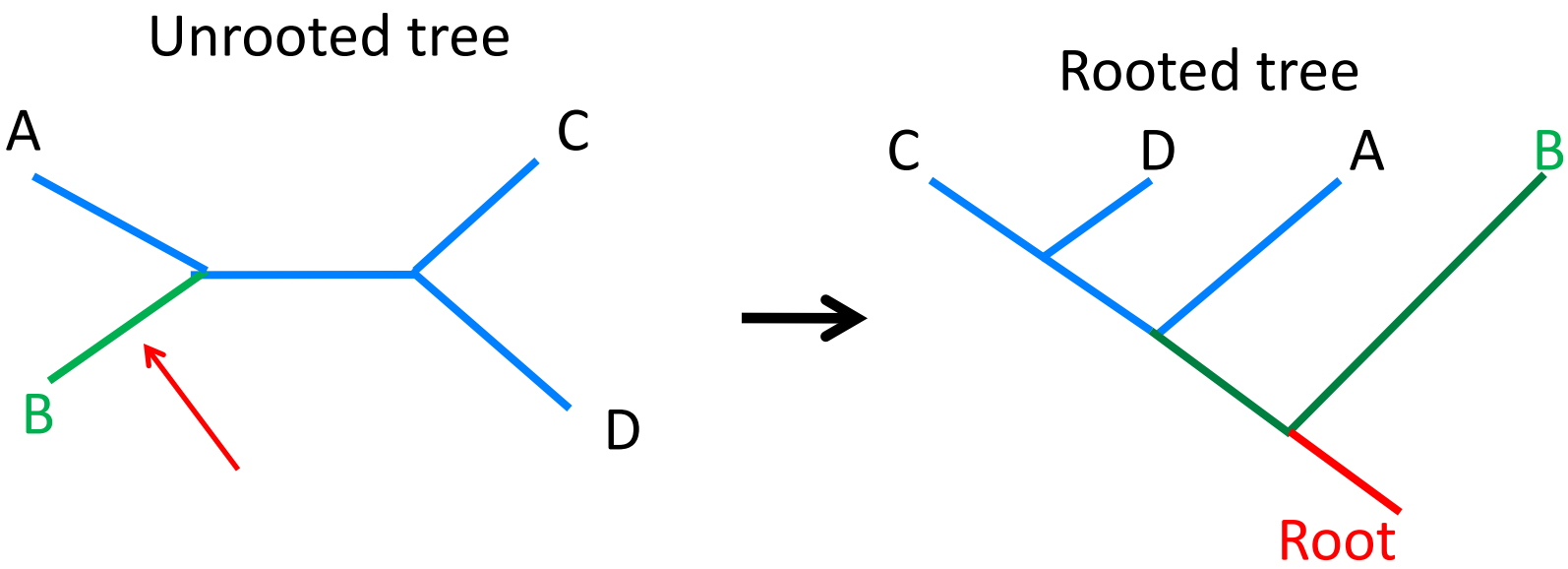


What do we get if we root the tree on the branch leading to B?

Unrooted tree

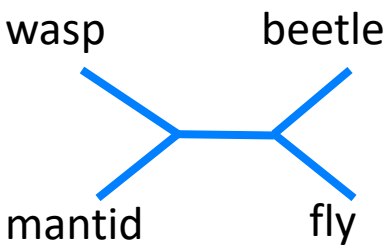


What do we get if we root the tree on the branch leading to B?

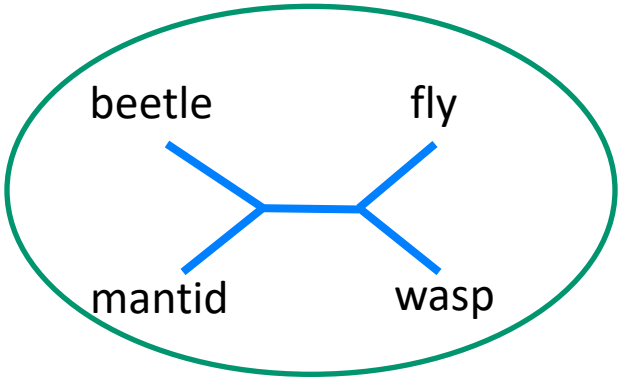


Among these four insects, which would be a reasonable outgroup?

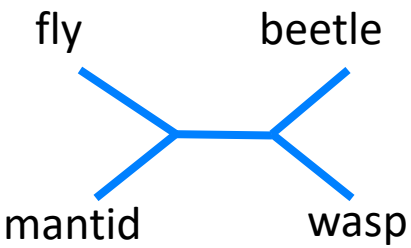
Tree #1



Tree #2



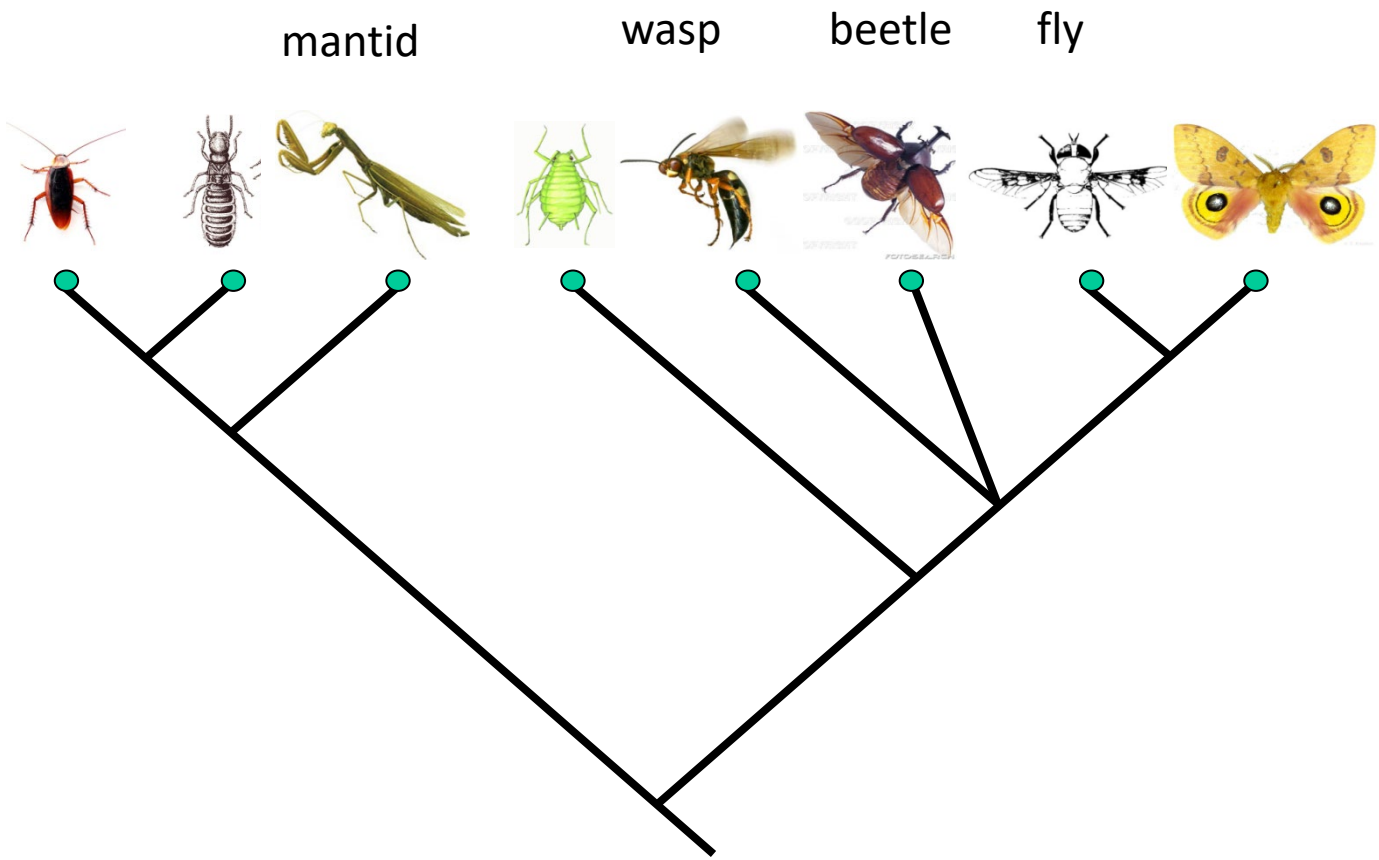
Tree #3



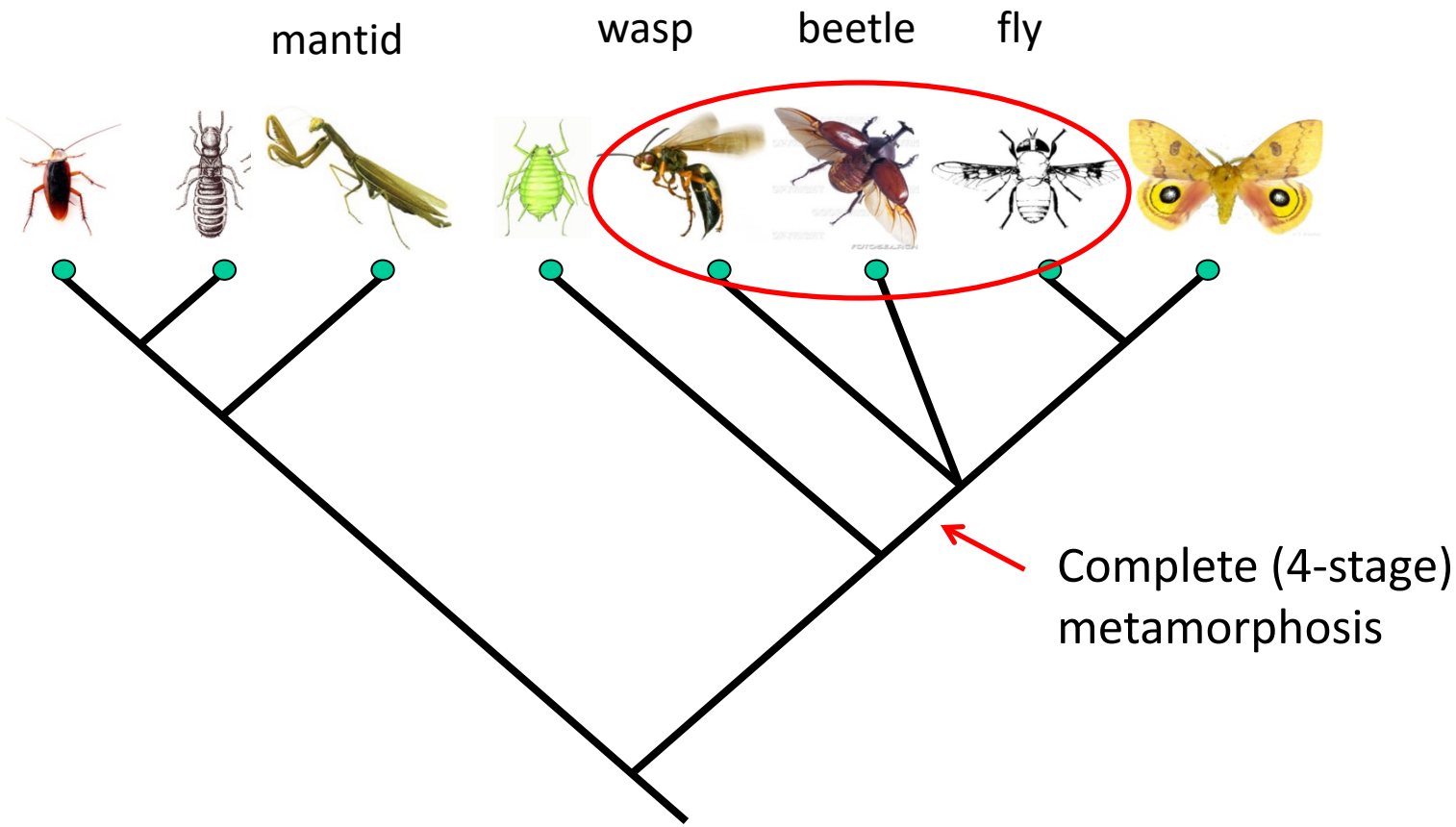
Char 3	2	1	2
Char 6	2	1	2
Char 7	1	2	2
Total	**5 + 3 = 8**	**4 + 3 = 7**	**6 + 3 = 9**

Plus three additional steps for each of the three parsimony-uninformative characters

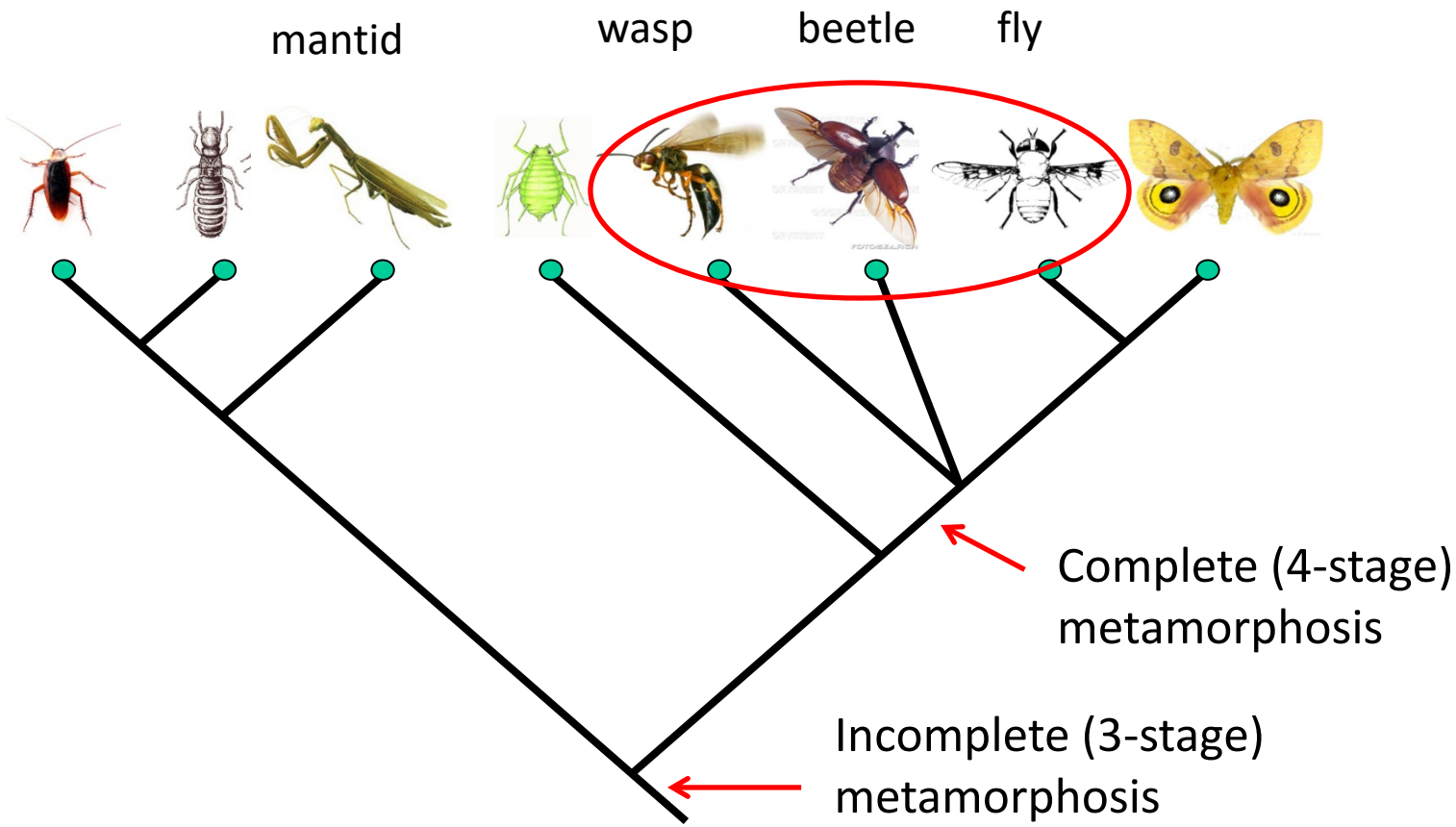
Based on external knowledge…



Based on external knowledge…

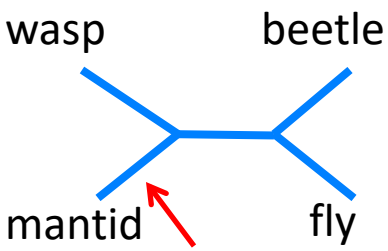


Based on external knowledge…

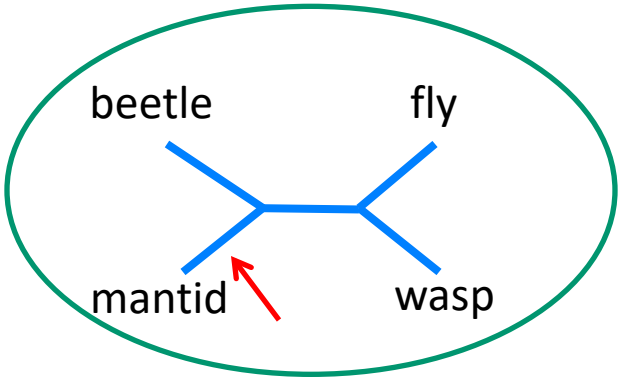


Among these four insects, which would be a reasonable outgroup?

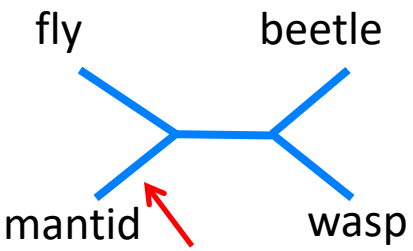
Tree #1



Tree #2



Tree #3

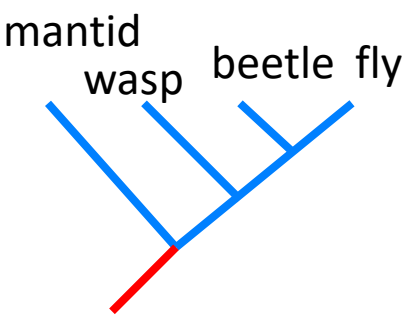


Char 3	2	1	2
Char 6	2	1	2
Char 7	1	2	2
Total	**5 + 3 = 8**	**4 + 3 = 7**	**6 + 3 = 9**

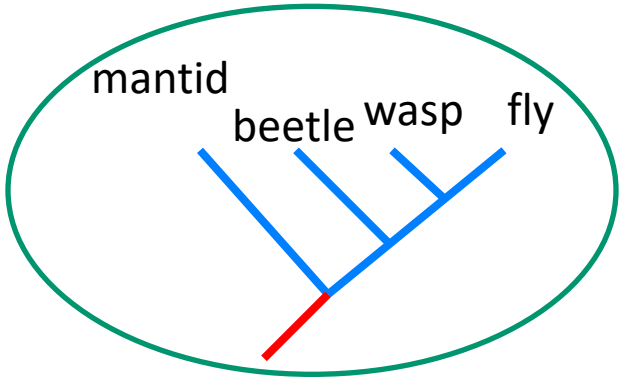
Plus three additional steps for each of the three parsimony-uninformative characters

Three rooted trees (mantid as outgroup)

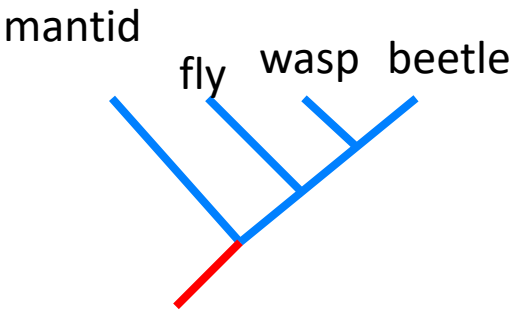
Tree #1



Tree #2



Tree #3



Char 3	2	1	2
Char 6	2	1	2
Char 7	1	2	2
Total	**5 + 3 = 8**	**4 +3 = 7**	**6 + 3 = 9**